DOCUMENT RESUME

ED 239 002 UD 023 305

AUTHOR Walberg, Herbert J.
TITLE Desegregation and Educational Productivity. Final Report.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Nov 82
NOTE 60p.; Paper submitted as one of a collection from the National Institute of Education Penel on the Effects of School Desegregation. For related documents, see UD 023 302-308. Some pages are cropped.
PUB TYPE Information Analyses (070) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Gains; Black Students; *Desegregation Effects; Effect Size; Elementary Secondary Education; Learning Processes; *Meta Analysis; Outcomes of Education; *Performance Factors; Productivity; Program Effectiveness; Program Evaluation; *Research Reports; School Desegregation; School Effectiveness; *Synthesis

ABSTRACT

This paper compares the effects of desegregation on black achievement with the effects of other factors in the process of school learning that have recently been synthesized. The first section of the paper discusses techniques and guidelines for research synthesis, including meta-analysis. The second section presents a summary of the statistical analyses of research reviews of the 1970s and a collection of meta-analyses of the 1980s, which reveal the consistently potent productivity factors in school learning and which further illustrate techniques for research synthesis. The third section assesses selection criteria for studies of school desegregation and achievement and compares the effects of desegregation--as revealed by three recent meta analyses--with the effects of the educational productivity factors. It is concluded that the amount and quality of instruction, constructive classroom morale, stimulation in the home environment, and other such productivity factors are more effective in increasing black achievement than is school desegregation. (CMG)

Desegregation and Educational Productivity

Herbert J. Walberg

University of Illinois at Chicago

Commissioned by the

National Institute of Education

November, 1982

Paper submitted as one of a collection from the National Institute of Education Panel on the Effects of School Desegregation

Running Head: Final Report

The purpose of the present paper is to analyze research on the impact of school desegregation on academic achievement. More specifically, the particular emphasis of this paper is the comparison of the effects of desegregation with those of other factors in the process of school learning that have been recently synthesized.

The paper is divided into three sections. The remainder of this first section discusses techniques and guidelines for research synthesis including meta-analysis. The second section presents a summary of the statistical analyses of research reviews of the 1970's and a collection of meta-analyses of the 1980's, which reveal the consistently potent productivity factors in school learning and which further illustrate techniques and guidelines for research synthesis. The third section assesses selection criteria for studies of school desegregation and achievement, and compares the effects of desegregation--as revealed by three recent meta-analyses--with the effects of the educational-productivity factors.

Research Synthesis

The present is an extraordinary time in the history of education because research syntheses are demonstrating the consistency of educational effects and are helping to put teaching and other determinants of learning on a sound scientific basis. Research synthesis is an attempt to apply scientific techniques and standards explicitly to the evaluation and summarization of research; it not only

statistically summarizes effects across studies but also provides detailed, replicable rationales and descriptions of literature searches, selection of studies, metrics of study effects, statistical procedures, and overall results as well as those that call for exception with respect to context or subjects by objective statistical criteria (Glass, 1977; Cooper & Rosenthal, 1980; Jackson, 1980; Walberg & Haertel, 1980; Glass, McGaw, & Smith, 1981; and Light & Pillemer, 1982). Qualitative insights may be usefully combined with quantitative synthesis (Light & Pillemer, 1982); and quantitative results from multiple reviews and syntheses of the same or different topics may be compiled and compared to estimate their relative magnitudes and consistencies (Walberg, 1982).

Research synthesis is not merely statistical analysis of studies. Jackson (1980) discusses six tasks comprising an integrative review or research synthesis: specifying the questions or hypotheses for investigation; selecting or sampling the studies for synthesis; coding or representing the characteristics of the primary studies; analyzing, or meta-analyzing (Glass, 1977) or statistically synthesizing the study effects; interpreting the results; and reporting the findings.

Although these tasks seem obviously necessary to encourage replication of reviews, Jackson found only 12 out of 87 recent reviews in prominent educational, psychological, and sociological journals that provided even

a cursory statement of methods. The basic idea behind much good advice in Jackson's paper is that the methods of review and synthesis should be explicit to enable other investigators to attempt to replicate the synthesis.

Explicit methods concerning quantitative synthesis, however, inevitably call for statistics, and two are most often employed--the vote count or box score, and the effect size (Glass, 1977). The vote count is easiest to calculate and explain to those who are unaccustomed to thinking statistically; it is simply the number of percentage of all studies that are positive, for example, in which the experimental exceeded control groups or the independent variable correlated positively with the dependent variable.

The effect size is the difference between the means of the experimental and control groups divided by the control group standard deviation; it measures the average superiority (or, inferiority, if negative) of the experimental relative to the control groups (for cases in which these statistics are unreported, Glass (1977) provides a number of alternate estimation formulas). If education had uniform ratio variables such as time and money as in economics or physical measures in natural sciences such as meters and kilograms, effect sizes would be unnecessary; it could be said, for example, that the experimental groups grew .42 comprehension units in reading history on average, and the control group grew .22 units without crude post hoc standardization for comparability required in meta-analysis.

Effect sizes permit a rough calibration of comparisons

across tests, contexts, subjects, and other characteristics of studies. The estimates, however, are affected by the variances in the groups, the reliabilities of the outcomes, the match of curriculum with outcome measures, and a host of other other factors, whose influences, in some cases, can be estimated specifically or generally. Although effect sizes are subject to distortions, many of which may counterbalance one another, they are the only means of comparing the size of effects in primary research that employs various outcome measures on non-uniform groups. They are likely to be necessary until an advanced theory and science of educational measurement develops ratio measures that are directly comparable across studies and populations.

Generalizability

The generality of the results of the synthesis can be divided into questions of extrapolation and interpolation: Do the synthesized results generalize to other populations and conditions, particularly to those that have not been studied or for whom the results are unpublished? And, do the results generalize across populations and conditions for which results are available? Extrapolation may be invalid beyond published studies because journal editors favor positive, significant studies. Smith (1980) estimates from several syntheses that mean effect sizes in unpublished work, mainly doctoral dissertations, are occasionaly larger but average about a third smaller than those in published studies.

Rosenthal (1980), on the other hand, shows that, given the great statistical significance of collections published studies, the probability of null effects being established by unpublished studies is minimal. Furthermore, both the low reliability of educational measures and low curricular validity (correspondance of what is taught and what is tested on outcome measures) diminish the estimates of relations between educational means and ends. Less than optimal reliability and validity, which leads to underestimates of effects, probably more than compensate for publication bias; but more empirical and analytic work is needed on these factors to determine their general and specific influences on synthesis results.

Interpolation

The interpolation problem can be readily solved by additional calculations. The most obvious questions in quantitative synthesis concern the overall percentage of positive results and their average magnitude. But the next questions should concern the consistency and magnitude of results across student and teacher characteristics, educational treatments and conditions, subject matters, study outcomes, and validity factors in the studies. These questions can be answered by calculating separate results for classifications or cross-classifications of effects.

The results may be compared by objective statistical tests (such as T, F, and regression weights in general linear models). They permit conclusions on such matters as

the overall effectiveness of treatments as well as their differential effectiveness on categories of students in various conditions on different outcomes. Notwithstanding the frequent claims by reviewers for differential effects on the basis of results of a few selected studies, most research syntheses yield results that are robust and roughly consistent across such categories. Such robustness is scientifically valuable because it indicates parsimonious, law-like findings; it is also educationally valuable because educators can apply robust findings more confidently and efficiently rather than using complicated, expensive procedures, tailor-made on unproven assumptions to special cases.

A number of useful methodological writings are available. Glass (1977) provides a concise introduction to statistical methods; and Glass, McGaw, and Smith's (1981) book presents a comprehensive treatment. Jackson (1980) and Cooper (1982) discuss tasks and criteria for integrative reviews and research syntheses. Light and Pillemer (1982) decribe methods for combining quantitative and qualitative methods. Walberg and Haertel (1980) present a collection of eight methodological papers by Cahen, Cooper, Hedges, Light, Rosenthal, Smith and others and thirty-five substantive papers mostly on educational topics. In forthcoming work, Larry Hedges of the University of Chicago and Barry McGaw of Murdoch University (Australia) offer firmer statistical and psychometric footings for quantitative synthesis. Important

guidelines for research synthesis that may be found in these works are further discussed and illustrated in the remaining sections.

Educational Productivity Factors

A Review of Reviews of Teaching Effects

The year 1980 marked a transitional period when investigators recognized the shortcomings of the traditional review and the advantages of more objective, explicit procedures for evaluating and summarizing research. Yet reviews still have a place, and much can be learned from them. Waxman and Walberg (1982) examined 19 reviews of teaching process-student outcome research published during a recent decade that critically reviewed at least three studies and two teaching constructs; they described their methods, compared their conclusions, synthesized them, and and pointed out the implications for future reviews, syntheses, and prior research.

The 19 reviews reflect the inexplicit, varied, and vague standards revealed by Jackson's (1980) analysis of 87 review articles in prominent educational, psychological, and sociological journals. None of the reviews, for example, described their search procedures, and only one stated explicit criteria for inclusion and exclusion of primary studies. Comparative analysis of the studies, moreover, revealed that the reviewers failed to search diligently enough for primary studies or to state the reasons for excluding large parts of the research evidence. Among the

five reviews that covered positive reinforcment such as praise and feedback in teaching, only six studies were covered in the most comprehensive review in contrast to the 39 listed in Lysakowski and Walberg's (1981) synthesis. Such arbitrary selection of small parts of the evidence, of course, leaves the reviews open to systematic bias and means that the reviews and their conclusions cannot be replicated in a strict sense because their methods are undescribed.

Although the reviews purported to be critical, their coverage of the 33 standard threats to methodological validity (Cook & Campbell, 1979) was spotty and haphazard. In 95.4 percent of the possible instances, the reviews ignored specific threats. External validity (interaction of teaching treatments with selection, setting, and history) was relatively well covered, perhaps reflecting the search and claims for aptitude-treatment interactions of the 1970's; but the serious problem of internal validity such as reverse and exogenous causes in correlational studies were almost wholly ignored. Indeed, there appeared an odd tendency to select correlational studies rather than experiments for review.

Despite these problems, however, a statistical tabulation of the conclusions of the reviews shows substantial and statistically-significant agreement that five broad teaching constructs--cognitive cues, motivational incentives, engagement, reinforcement, and management and climate--are positively associated with student learning

outcomes (see Table 1). These tabulations, moreover, are in close agreement with quantitiative syntheses of large, systematic collections of primary studies discussed in a subsequent section.

Insert Table 1 about here

Current Research Syntheses

To characterize quantitative syntheses of educational research completed since 1979, sixteen were found in 1982 by scanning publications of the American Educational Research Association and writing to the members of "the invisible college" of about 100 scholars that meet annually to present and discuss research on teaching. A more systematic search in late 1982 using Dissertation Abstracts, Social Science Citation Index, Education Index, computer retrieval, and references in recent publications indicates that these syntheses plus those discussed in subsequent sections of this chapter represent about three-fourths of those completed in education thusfar in the 1980s. (An analysis of a more complete corpus is underway by the present author and colleagues, but the increasing number of syntheses makes exhaustive coverage an elusive goal.)

Table 2 suggests a number of instructive points for both educational practice and research synthesis. It provides, for example, an empirical answer to the coincidence of vote counts and effect sizes. Every mean

effect size that was positive also had a vote count greater than 50 percent; every negative effect size had a vote count less than 50 percent. Thus, as may be expected from normal distributions, consistently positive findings will yield positive average results (the next section shows that much of the variance in effects can be predicted by regression from counts). The likely explanation for the uniform association is that strong causes produce results consistent in sign. Indeed, the only cases in which the association can be reversed are skewed distributions in which a few very strong positive results are sufficient to pull the mean above zero from a cluster of small effects, more than half of which are negative (or vice versa).

Insert Table 2 about here

The first two syntheses grouped under Teaching Stategies in Table 2 show fairly close agreement with respect to the consistency of cooperative learning. Johnson and others (1981) categorized their results by comparisons of four treatment variations (cooperative, competitive, group competitive, and individualistic), whereas Slavin (1980) categorized his results by outcomes. Cooperative learning obviously produces superior results; but it would be useful if journal editors would allow research synthesists space to report average results by more standard classifications of independent and dependent variables and

study conditions to facilitate comparisions of replicated syntheses such as these two.

The next two syntheses raise important, unresolved methodological questions. Becker and Gersten's (1982) synthesis indicated a small average effect of direct instruction in several sites, but all effect sizes came from the same study. Although teachers in the various sites may have been independent actors, methodological bias can make the effects non-independent from a statistical point of view, and independent replications by different investigators would be in order to a provide a more definitive answer. Pflaum and others (1980) found no average superiority of different reading methods but a substantial advantage in learning outcomes of experimental over control groups no matter what the reading method employed. Although Hawthorne effects could be discounted by the synthesis, the increased energy and attention devoted to tasks by teachers in experimental groups rather than putative treatments themselves may partly account for superior results in teaching-methods and other educational studies.

Table 2 includes two rough replications that indicate substantial agreement in results despite large variations in study search, selection, and numbers. Hansford and Hattie's (1982) and Findley and Cooper's (1981) syntheses of correlations of self-concept and locus of control with achievement and performance differ only slightly in the second decimal place in both the vote counts and average

correlations. Carlberg and Kavale's (1980) and Ottenbacher and Cooper's (1981) syntheses agree that the effects of mainstreaming (federally-encouraged efforts in the United States to mix regular and cognitively, emotionally, and physically handicapped children in the same classes) are inconsistent and probably near zero.

Two syntheses show curvilinear effects of independent variables on educational outcomes. Smith and Glass (1980) found that the benefits of reduced class size are larger at the smaller ranges of one to 10 members than they are at higher ranges; for example, the measureable cognitive and affective outcome differences between classes of 20 and 60 appear trivial. Similarly, Williams and others (1982) found decreasing achievement with departures from 10 weekly hours of leisure-time television viewing such that estimated differences in achievement between children who watch about 30 hours--an average number--and 60--a large amount--are miniscule.

Other effects are summarized in the table, and the reader is referred to the original syntheses for details that are not discussed here. Overall the results indicate a large range of effects, which, if replicated in further primary research and syntheses, would have fairly definite implications for choosing policies and practices that seem likely to have consequential effects on raising educational outcomes.

The Michigan Program

Chen-Lin and James Kulik lead a vigorous group of research synthesists at the University of Michigan, which included Peter Cohen, now of Dartmouth. The group has been unusually productive of high-quality syntheses first in higher education and later in secondary-school research. Personal communications with the group reveal that their team approach, much like that described by Shulman and Tamir (1973) in the Second Handbook of Research on Teaching, accounts in part for the quantity and quality of work.

James Kulik kindly prepared Table 3 according to the present author's specifications. It shows the results of eleven syntheses completed by the Michigan group by the end of 1981. Like the sixteen syntheses by other investigators discussed in the last section, those in Table 3 show a number of consistent moderate to large effects that can help to put high school and college teaching on a firm scientific basis.

Insert Table 3 about here

Kulik's results also permit an estimate of the mean size of effects from vote counts. The regression equation, ES = -.403 + .008 (% Positive), accounts for 76 percent of the variance in the effect sizes. The corresponding equation for the syntheses in Table 2 for which both indexes are available, ES = -.761 + .015 (%), accounts for 59 percent of the effect-size variance (the correlational

results assume both causality and a one-unit increase in the independent variable). Both equations forecast near zero effect sizes for vote counts of 50 percent; but the higher slope for the results in Table 2 forecast larger effects than do the Michigan data; at vote counts of 75 percent, for example, the respective forecasts are .36 and .20. Thus the size of the regression slope is unstable across samples, and more intensive analyses of the complete corpus of syntheses are in order.

The two data sets also permit separate empirical estimates of the distributions of vote counts and effects. The mean (and standard deviations) of Michigan and other estimates of the vote counts are respectively 67 and 64 (and 19 and 16); the mean effects are respectively .17 and .22 (and .19 and .31). Assuming normal distributions of effects, empirical norms for vote counts and effect sizes can be set forth on the basis of the averages of these statistics; for example, the middle two-thirds of the effects in the recent educational research sampled range from about -.05 to .45. It could be said that effect sizes of .20 are average, and those above .45 are large and exceed about 84 percent of those typically found in educational research. Similarly, vote counts of 67 and 85 percent might be provisionally taken as average and large. These norms are, of course, very rough and preliminary, but they are based empirical results rather than opinion and may be useful in gauging present and future results until larger normative samples are analyzed.

Syntheses of Bivariate Productivity Studies

A group at the University of Illinois at Chicago has concentrated on synthesizing research on nine theoretical constructs that appear to have consistent causal influences on academic learning: student age or developmental level, ability (including prior achievement), and motivation; amount and quality of instruction; the psychological environments of the class, home, and peer group outside school; and exposure to the mass media (Walberg, 1981). The group first collected available vote counts and effect sizes in the review literature of the 1970's and then conducted more systematic syntheses directly on the nine factors. This section summarizes both efforts.

Synthesis of reviews of the 1970's. Walberg, Schiller, and Haertel (1979) collected reviews published from 1969 to 1979 on the effects of instruction and related factors on cognitive, affective, and behavioral learning in research conducted in elementary, secondary, and college classes and indexed in standard sources. The vote counts for the corpus of reviews are shown in Table 4.

Insert Table 4 about here

The vote counts should be cautiously interpreted because not only may journal editors more often select studies with positive results but also reviewers may select

positive published studies for summarization. Neither editors nor reviewers ordinarily state explicit policies on these important points. Subsequent, more systematic syntheses, nonetheless, have generally supported traditional reviews; and it would be wasteful to ignore the labors of the last decade of effort, even though it may only be considered a starting point for subsequent work.

Notwithstanding the possible double bias in the vote counts (see earlier sections on counter-biases), the results in Table 4 are impressive. A majority of the variables in the table were positively associated with learning; in 48 or 68 percent of the 71 tabulations, 80 percent or more of the comparisons or correlations are positive. Although all of the variables are candidates for synthesis using systematic search, selection, evaluation, and summarization procedures, it appears that the 1970's produced reasonably consistent findings that are likely to be confirmed by more comprehensive and explicit methods of the present decade.

Syntheses of Productivity Factors. The Chicago group also carried out syntheses of the nine factors using methods discussed in previous sections of this chapter. The National Institute of Education supported the syntheses of learning research in ordinary classes, grades kindergarten through twelve. A separate grant from the National Science Foundation on science learning, grades 6 through 12, permitted more exhaustive, intensive search for unpublished work and an advisory group of science educators and research

methodologists as well as a semi-independent replication of the results for several of the factors. A summary of the findings is shown in Table 5.

Insert Table 5 about here

All of the effect sizes (including mean contrasts and correlations) are in the expected direction. The mean effects for the two samples of studies are similar in magnitude, which suggests generality or robustness of effects across more and less intensive methods of synthesis. In particular, the syntheses of quality of instruction including cues, participation, and reinforcement of about 1.0 and .8 in general grades K-12 and in science grades 6-12 support the conclusions of the 19 reviews discussed in a previous section (see also Table 1). Despite these corroborations of findings, of course, independent replications of the syntheses as well as new and probing experimental studies are needed.

Syntheses of Multivariate Studies

The Chicago group also conducted multivariate analyses of the productivity factors in samples of from two to three thousand 13- and 17-year-old students who participated in the mathematics, social studies, and science parts of the National Assessment of Educational Progress (see, for example, Walberg, Pascarella, Haertel, Junker, and Boulanger, 1981, 1982). These survey analyses complement

small-scale correlational and experimental studies in providing on representive national samples data on fairly comprehensive sets of the productivity factors, each of which may be statistically controlled for the others in multiple regressions of achievement and subject-matter interest.

Such analyses allow a simultaneous assessment of qualities and amounts of instruction and the other factors in the production of learning. Since the factor levels are reported as experienced by individual students, the analysis are sensitive to micro-variations in the multiple environments of the school, peer-group, home, and mass media to which each student is exposed.

Although the sets of variables available in the National Assessment can be used to assess possible exogenous causes because they are measured and can be statistically controlled in regression equations, the measures are cross-sectional for individuals. Therefore, they cannot effectively rule out reverse causation such as learning as a cause of motivation and more stimulating teaching. Another shortcoming of the data is that parental socioeconomic status serves as a proxy for ability and prior achievement.

As pointed out above, nonetheless, the strengths of the National Assessment data complement those of small-scale bivariate studies that typically control for only one or two of the factors. If syntheses of both data sources point in the same direction, then more confidence can be placed in the conclusions.

Table 6 shows that the factors, when controlled for one another, are suprisingly consistent in sign, significance, and magnituide across subject matters, ages, operational measures of the factors, and independent national samples. The median standardized regression weights and squared multiple correlations, shown in the last row, reveal the small to moderate effects of the factors when controlled for one another and sizable amounts of variance accounted for even without ability and prior achievement measures.

Insert Table 6 about here

Syntheses of Open Education Research

Open education is an elusive concept, now dismissed by many educators, but one that research synthesis now illuminates. The history of efforts to synthesize its effects is instructive about: the dangers of basing conclusions, policies, and practices on single studies; replication and improved methods of syntheses, and a shortcoming of much of the research discussed above that employs grades and standardized achievement as the sole outcomes of teaching.

From the start, open educators tried to encourage educational outcomes that reflect school-board goals such as cooperation, critical thinking, self reliance, constructive learning attitudes, life-long learning, and other goals that evaluators seldom measure. Raven's (1981) summary of

surveys in Western countries including England and the United States shows that educators, parents, and students rank these goals far above standardized test achievement and grades.

A synthesis of the relation of conventionally-measured educational outcomes and adult success, moreover, shows their slight association (Samson and others, 1982). Thirty-three post-1949 studies of physicians, engineers, civil servants, teachers, students in general, and other groups show a mean correlation of .155 of these educational outcomes with success indicators such as income, self-rated happiness, work performance and output indexes, and self-, peer-, and supervisor-ratings of occupational effectiveness. These results should challenge educators and researchers to seek a balance between continuing motivation and skills to learn and perform well on new tasks as an individual or group member on one hand and mastery of teacher-chosen, textbook knowledge that may soon be obsolete or forgotten on the other.

Perhaps since Socrates, however, arguments over student-centered and teacher-centered education have remained so polarized, polemical, and pervasive that educators find it difficult to stand firmly on the high middle ground of balanced, joint, or cooperative determination of the goals, means, and evaluation of learning. Progressive education, the Dalton and Winnetka plans, team teaching, the ungraded school, and other

innovations in this century held forth this ideal but gravitated toward authoritarian teaching or permisiveness and could not be sustained. Although open education, too, faded from view, it was more carefully researched; and syntheses of it may help prepare educators for evaluating future efforts.

Three Syntheses of Open Education. Horwitz (1979) first synthesized about 200 comparative studies of open and traditional education by tabulating vote counts by outcome category. Although many studies yielded non-significant or mixed results especially with respect to academic achievement, self concept, anxiety, adjustment, and locus of control, more positive results were found in open education on attitudes toward school, creativity, independence, curiosity, and cooperation.

Peterson (1979) calculated effect sizes for the 45 published studies. She found about -.1 or slightly inferior effects of open education on reading and mathematics achievement; .1 to .2 effects on creativity, attitudes toward school, and curiosity; and .3 to .5 effects on independence and attitudes toward the teacher.

Hedges, Giaconia, and Gage (1981) synthesized 153 studies including 90 dissertations using an adjustment of Glass's effect-size estimator which is slightly biased especially in small samples. The average effect was near zero for achievement, locus of control, self concept, and anxiety; about .2 for adjustment, attitude towards school

and teacher, curiosity, and general mental ability; and about .3 for cooperativeness, creativity, and independence.

Despite the differences in study selection and synthesis methods, the three studies converge roughly on the same plausible conclusion: students in open classes do slightly or no worse in standardized achievement and slightly to substantially better on several outcomes that educators, parents, and students hold to be of great value. Unfortunately, the negative conclusion of Bennett's (1976) single study--prefaced by a prominent psychologist, published by Harvard University Press, publicized by the New York Times and media and experts that take that newspaper as their source--probably sounded the death knell of open education, even though the conclusion of the study was later retracted (Aitkin, Bennett, & Hesketh, 1981) because of obvious statistical flaws in the original analysis (Aitkin, Anderson, & Hinde, 1981).

Components of Open Education. Giaconia and Hedges (1982) took another recent and constructive step in the synthesis of open education research. From the prior effect-size synthesis, they identified the studies with the largest positive and negative effects on several outcomes to differentiate more and less effective program features. They found that programs that are more effective in producing the non-achievement outcomes--attitude, creativity, and self concept--sacrificed academic achievement on standardized measures.

These programs were characterized by emphasis on the

role of the child in learning, use of diagnostic rather than norm-referenced evaluation, individualized instruction, and manipulative materials but not three other components sometimes thought essential to open programs--multi-age grouping, open space, and team teaching. Giaconia and Hedges speculate that children in the most extreme open programs may do somewhat less well on conventional achievement tests because they have little experience with them. At any rate, it appears from the two most comprehensive syntheses of effects that open classes on average enhance several non-standard outcomes without detracting from academic achievement unless they are radically extreme.

Synthesis of Instructional Theories

To specify the productivity factors in further theoretical and operational and detail provide a more explicit framework for future primary research and synthesis, Haertel, Walberg, and Weinstein (1983) compared eight contemporary psychological models of educational performance. Each of the first four factors in Table 7--student ability and motivation, and quality and quantity of instruction--may be essential or necessary but insufficient by itself for classroom learning (age and developmental level are omitted because they are unspecified in the models).

Insert Table 7 about here

The other four factors in Table 7 are less clear: although they consistently predict outcomes, they may support or substitute for classroom learning. At any rate, it would seem useful to include all factors in future primary research to rule out exogenous causes and increase statistical precision of estimates of the effects of the essential and the other factors.

Table 7 shows that, among the constructs, ability and quantity of instruction are widely and relatively richly specified among the models. Explicit theoretical treatments of motivation and quantity of instruction, however, are largely confined to the Carroll tradition represented in the first four models; and the remaining factors are largely neglected.

The table poses empirically-researchable theoretical questions; the tension between theoretical parsimony and operational detail, for example, suggests several: Can the first four constructs mediate the causal influences of the last four? Would assessments of Glaser's five student-entry behaviours allow more efficient instructional prescriptions, than would, say, Carroll's, Bloom's, or Bennett's more general and more parsimonious ability subconstructs? Would less numerous subconstructs than Gagne's eight instructional qualities and Harnischfeger and Wiley's seven time categories suffice?

The theoretical formulation of educational performance

models of the past two decades since the Carroll and Bruner papers has made rapid strides. The models are explicit enough to be tested in ordinary classroom settings by experimental methods and production functions. Future empirical research and syntheses that are more comprehensive and better connected operationally to these multiple theoretical formulations should help reach a greater degree of theoretical and empirical consensus as well as more effective educational practice.

Desegregation and Educational Productivity

As the previous section has shown, sufficient empirical and theoretical syntheses have accumulated during the past five years to point more definitively than ever before to the proximal, alterable factors that affect educational achievement. Nearly all the research has been carried out in natural settings such as homes and schools, and most of it shows generalizability across student characteristics, subjects, and research methods, including randomized assignment to experimental treatments.

The large average magnitude and consistency of many of these productive factors justly provides a substantial amount of confidence about how educational achievement may be raised. Since many of the factors and techniques have already been extensively employed in ordinary schools and found successful, inexpensive, and non-controversial, it appears that educational achievement might be increased

substantially by implementing a selection of the most productive of the factors, say, those with effect sizes above .3, more extensively and intensively. The purpose of this section is to compare the consistency and magnitude of such factors to the effects of school desegregation, as revealed by three recent meta-analyses--Krol (1978), Crain and Mahard (1982), and my statistical summary of the studies meeting the selection criteria of the National Institute of Education (NIE) panel of scholars.

Selection Criteria

Aside from the inclusion of data only on black students in all three meta-analyses, Krol (1978, p. 16), Crain and Mahard (1982, p. 6) and the NIE panel (Schneider, Note 1) varied considerably in explicit criteria for study selection. Krol, for example, excluded studies that lacked achievement measures before and after desegregation and those that lacked sufficient statistics to calculate effect sizes (pp. 83-84). Excluding studies without pretests turns out to be a reasonable decision because Wortman's (Note 2) research shows desegregated groups are on average advantaged on achievement before desegregation. Thus apparent posttest advantages of desegregation are in part attributable to pre-existing differences, and pretest adjustment is required for valid estimation of desegregation effects.

Crain and Mahard (1982) excluded "excluded a large number of papers, many of which compared students in racially segregated and racially mixed schools, but gave no

indication that a formal desegregation plan had been adopted" (p. 6). Because they included studies that employed ability (in contrast to educational achievement) as a dependent variable and conducted a more recent and exhaustive search, they used 93 studies for analysis in contrast to Krol's 55 (see Tables 8 and 9).

Insert Tables 8 and 9 about here

The NIE panal employed a number of stringent criteria for study rejection including the following: non-empirical and summary reports; studies done outside the U. S. and geographically non-specific; those that combined or compared ethnic groups, lacked contemporaneous-control or pre-desegregation data, or analyzed heterogenously desegregated groups; those with more than 35 percent attrition, majority-black desegregated conditions, varied exposure to desegregation, and non-comparable groups; those with unknown sampling procedures, cross-sectional data, or non-comparable samples at each observation point; those with unreliable or unstandardized instruments, unknown test content or instruments, unknown test administration dates, ability tests as dependent variables, and non-equivalent pre-tests and post-tests; and insufficient statistics (Schneider, Note 1). Application of these exclusion criteria (Wortman, Note 2) resulted in 19 "acceptable studies."

Thus, all three data sets are similar in including only

studies of black achievment. The differ chiefly in that Krol and the NIE panel, unlike Crain and Mahard (1982), exclude ability tests, and the NIE employed stringent methodological criteria that resulted in a selection of studies only 19 percent as large as Crain and Mahard's set (see Table 8).

The NIE panel may be right in specifying stringent selection criteria from one viewpoint: the conclusions of review articles are usually based upon methodologically acceptable studies. But, as Glass, McGaw, and Smith (1982, p. 226) point out, excluding studies by implicit or explicit selection criteria can convert empirical questions of research methodology to a priori assumptions. Excluding studies without pretests, for example, may exclude randomized experiments--possibly the best design in certain respects for probing causality and avoiding untenable covariance assumptions.

If it were to be found that randomized post-test only designs yielded the same results as pre-test-post-test quasi-experiments, then greater confidence could be placed in the results than the results of either design by themseves, since the two designs are subject to different threats to methodological validity (Cook & Campbell, 1979). Because, for example, the findings on instructional research are generally robust and consistent across study features such as research methods and student characteristics, substantial confidence can be placed in their results.

Moreover, excluding studies on policy or substantive criteria may be useful to lighten the effort or to narrow research questions; but exclusion also restricts the inferences and comparisons that can be made and the policies that may be implied. In the Krol and NIE selections, for example, it will not be possible to determine whether desegregation has a different impact on achievement than it does on ability or other educational outcomes such as creativity, critical thinking, interest in further learning, and social perceptiveness. In none of the three sets of studies, moreover, will it be possible to compare the effects of desegregation on Asian, black, Hispanic, and white students. At least for some parents, educators, policy makers, researchers, and others, it would be useful to have reliable information on these and other points.

None of this is to argue that all studies should be summarized in one overall vote count or mean effect size. Although that statistic and its significance are of interest, characteristics of the studies such as Cook and Campbell's (1979) 33 threats to methodological validity, student characteristics such as ethnicity and grade level, and conditions of desegregations such as voluntary and mandatory plans should be categorized, coded, and tested for statistical significance with studies as the units to afford independence as assumed in statistical inference. (If desegregation is working generally well according to a study, then students in different grades within the study

are likely do well, and their performance is correlated and not statistically independent; similarly, if students are doing poorly in another study, different grades lack independence; therefore the means for studies, not for grade levels or other units, must be taken as the units for meta-analysis or each comparison in a study must be weighted inversely to the number of comparisons in the study. Another reason for using study means or weighting is to insure that each study is given an equal weighting of one, not a weighting based on the arbitrary number of comparisons the investigator happened to make.)

Synthesis of Three Meta-analyses

Tables 8 and 9 show what can be validly extracted as the chief findings from the three meta-analysis. Table 8 shows that three estimates of percent-positive studies vary between 61 and 64 percent. These percentages are in surpirsingly close agreement considering the widely different selection criteria and numbers of studies in the three syntheses.

Table 9 shows that the statistical significance cannot be determined in two cases because the percentage of positive comparisons rather than studies are reported; and, in the NIE case, the sign test based on the number of studies is insignificant. By the norms of recent syntheses of productivity factors discussed in previous sections, the percentage magnitudes are neither large (85 percent) nor average (67 percent). The statistical significance of the

percentages cannot be determined in the two previous syntheses previously reported and is insignificant in the case of the NIE selection.

The statistical significance of the effect sizes are mixed: indeterminate for Krol, because of comparison weighting; significant for Crain and Mahard; and not significant for the set of studies acceptable to the NIE panel. In none of the three cases was the magnitude of the effect large (.45) or average (.20). (Crain and Mahard's significant finding of higher effects in kindergarten and first grade are unsupported by Krol and reversed in analyses by Wortman (Note 2); and their randomized-longitudinal effect is insignificant with study as the unit. Thus, their overall average study-weighted effect size is reported in Table 8.)

The results from the three meta-analyses suggest that the vote counts fail with some uncertainty to reach conventional levels of statistical significance. By normative standards of recent syntheses of other educational factors, they clearly fail with respect to percentage results. The effect sizes as a set are indeterminate with respect to significance and certainly fail to reach criterion levels with respect to normative magnitude.

Conclusion

New techniques of research syntheses show a number of potent factors for improving educational achievement that have proven to be consistently effective in a wide variety

of experimental and educational conditions. These include the amount and quality of instruction, constructive classroom morale, and stimulation in the home environment. It is in our national economic, social, and political interest to implement these factors more deeply and widely for all children (Walberg, 1983). In this effort, school desegregation does not appear to prove promising in the size or consistency of its effects on learning of black students.

Reference Notes

1. Schneider, J. M. Personal communications. August 16, 1982; November 4, 1982.
2. Wortman, P. Personal communications. August 28, 1982; November 10, 12, 1982.

Table 1

Conclusions of 19 Reviews and 2 Quantitative Syntheses of Research on Teaching

	Stimulation		Engagement	Reinforcement	Management and Climate
	Cognitive Cues	Motivational Incentives			
Number of Reviews Covering Construct	19	5	10	13	15
Number of Reviews Concluding Relation to Learning is Positive	17	5	10	9.5	13.5
Probability of an Even Split	.01	.10	.01	.10	.01
Mean Effect Sizes from Quantitative Synthesis	1.28		.88	.94	1.17
Probability of Evidence Assuming Zero Population Effect	.01			.01	

Table 2

Selected Post-1979 Quantitative Syntheses

Author	Number of Studies	Independent and Dependent Variables	Mean Correlation or Effect	Percent Positive	Comments
Teaching Strategies					
Johnson, Maruyama, Johnson, Nelson, and Skon (1981)	122	Effects of cooperation, intergroup and interpersonal competition, and individual goal efforts on achievement and productivity	.00	54	Cooperative vs. group competitive
			.78	76	Cooperative vs. competitive
			.37	68	Group competitive vs. cooperative
			.76	83	Cooperative vs. individualistic
			.59	81	Group competitive vs. individualisti
			.03	47	Competitive vs. individualistic
Slavin (1980)	28	Effects of educational programs for cooperative learning		81	Curriculum-specific tests
				78	Standardized tests
				95	Race relations
				65	Mutual concern
Becker & Gersten (1982))	1	Effects of Direct Instruction Follow Through on later achievement (7 sites on 2 occasions, fifth and sixth grades)	.23	--	Effects larger for mathematics problem solving and for fifth grade
Pflaum, Walberg, Karagianes, and Rasher (1980)	96	Effects of different methods teaching reading on learning	.60	76	Although Hawthorne effects could be discounted, experimental groups generally did substantially better than controls; sound-symbol blendir was one standard deviation higher than other treatments.

Author	Number of Studies	Independent and Dependent Variables	Mean Correlation or Effect	Percent Positive	Comments
Teaching Skills					
Luiten, Ames, and Anderson (1980)	135	Effects of advance organizers on learning and retention	.23	--	Effects larger on 20+ days retention higher achievers, college students, and when presented aurally
Redfield and Rousseau (1981)	20	Effects of higher and lower cognitive questions	.73	--	Higher questioning effects greater training than in skills study and in more valid studies
Wilkinson (1980)	14	Effects of praise on achievement	.08	63	Praise slightly more effective for lower socioeconomic groups, primary grades, and in mathematics
Other Studies					
Dutcher (1981)	47	Effects of microteaching lessons on teaching performance of secondary and elementary education students	.84 .56 .46 .35	--	Secondary specific skills Secondary questioning skills Elementary specific skills Elementary questioning skills
Colosimo (1981)	24	Effects of practice and beginning teaching on self attitudes	-.29	48	Initial experience associated with greater authoritarianism and self doubt; inner-city experience more negative
Findley and Cooper (1981)	98	Correlations of locus of control and achievement	.18	79	Correlations higher among males; for adolescents in constrast to children and adult groups; for specific control measures; and for objective achievement

Author	Number of Studies	Independent and Dependent Variables	Mean Correlation or Effect	Percent Positive	Comments
Hansford and Hattie (1982)	128	Correlation of self-concept and achievement/performance	.21	84	Correlations higher for high school students in contrast to elementary and college; higher ability students; specific rather than global self-concept; and verbal achievement measures
Carlsburg and Ravale (1980)	50	Effects of special versus regular classes	-.12	--	Effects positive for learning disabled and behavior disordered and negative for slow learners and mentally retarded
Otterbacher and Cooper (1981)	43	Effects of class placement of mentally retarded students on social adjustments	.05 -.07	61 46	Special class vs. regular class Special class vs. resource class
Smith and Glass (1980)	59	Effects of class size on attitudes, climate, and instruction	.49	--	In contrast to small mean effect of .01 for achievement, moderate effect observed, which were larger on teacher than students, younger students, and for studies before 1969
Williams, Haertel, Haertel, and Walberg (1982)	23	Correlations of leisure time television and achievement	-.05	34	Effects negative at ratio of less than 5 or greater than 15 hours per week and stronger for girls and higher ability groups
Willson and Putnam (1982)	32	Effects of pretests on outcomes	.17	57	Effects greater for cognitive and personality outcomes, for treatment lasting between 2 and 30 days, and for randomized studies

Table 3

Major Results from Quantitative Syntheses Conducted at the University of Michigan's Center for Research on Learning and Teaching

Report	Independent Variable	Dependent Variable	Studies		Effect Size		Comments
			Number	% Positive	Mean	SD	
Bangert, J. Kulik, & C. Kulik (1981)	Individualized vs. conventional secondary teaching	Achievement on final examination	49	65	0.10	0.36	
		Attitude toward subject matter	14	64	0.14	0.27	
Cohen (1980)	Midsemester rating feedback to teacher vs. no feedback	Change on final ratings	22	91	0.38	0.41	Effects were greater when teachers received consulting help along with rating feedback.
Cohen (1981)	Class rating of instructor quality	Class achievement on final examination	67	88	0.43	0.13	Correlations were higher when teachers were faculty (not teaching assistants), when all tests were graded by a common grader, and when students rated teachers after receiving grades.
Cohen, Ebeling, & J. Kulik (1981)	Visual-based vs. conventional college teaching	Achievement on final examination	65	97	0.19	0.41	Achievement effects were stronger in more recent studies, in studies from universities, & when different teachers taught visual-based & control classes.
		Student rating of course quality	16	38	-0.06	0.66	
		Course completion	10	30	-0.05	0.23	

Table (Continued)

Report	Independent Variable	Dependent Variable	Studies		Effect Size		Comments
			Number	% Positive	Mean	SD	
J. Kulik, Cohen, & Ebeling (1980)	Programmed vs. conventional college teaching	Achievement on final examination	56	71	0.24	0.52	Achievement effects were stronger in more recent studies.
		Course completion	9	61	-0.08	0.27	
J. Kulik, C. Kulik, & Cohen (1979a)	Personalized System of Instruction vs. conventional college teaching	Achievement on final examination	81	94	0.49	0.33	Achievement effects differed by subject and were stronger when different teachers taught PSI and control classes, and when control classes contained PSI features.
		Course completion	27	37	-0.10	0.30	
		Rating of course quality	11	91	0.46	0.69	
J. Kulik, C. Kulik, & Cohen (1979b)	Audio-tutorial vs. conventional college teaching	Achievement on final examination	42	69	0.20	0.43	Achievement effects were stronger in studies found in journals.
		Course completion	22	52	-0.10	0.37	
		Rating of course quality	6	50	0.12	0.52	
J. Kulik, C. Kulik, & Cohen (1980)	Computer-based vs. conventional college teaching	Achievement on final examination	54	69	0.25	0.61	Achievement effects were stronger when different teachers taught computer-based and control classes.
		Course completion	13	46	0.01	0.30	
		Rating of course quality	11	73	0.24	0.52	

Table 4

A Selective Summary of a Decade of Educational Research

rch Topics	No. of Results	Percent Positive
on learning	23	95.4
vative curricula on:		
novative learning	45	97.8
ditional learning	14	85.7
ller classes on learning:		
e-1954 studies	53	66.0
e-1954 better studies	19	84.2
st-1954 studies	11	72.7
ll Comparisons	691	60.0
avioral instruction on learning	52	98.1
onal systems of instruction on learning	103	95.2
nery learning	90	96.7
dent-vs. instructor-led discussion on:		
chievement	10	100.0
ttitude	11	100.0
tual vs. conceptual questions on achievement	4	100.0
cific teaching traits on achievement:		
Clarity	7	100.0
Flexibility	4	100.0
Enthusiasm	5	100.0
Task-orientation	7	85.7
Use of student ideas	8	87.5
Indirectness	6	83.3
Structuring	9	100.0
Sparing criticism	17	70.6

Table 4 (Continued)

Research Topics	No. of Results	Percent Positive
Psychological incentives and engagement		
Teacher cues to student	10	100.0
Teacher reinforcement of student	16	87.5
Teacher engagement of class in lesson	6	100.0
Individual student engagement in lesson	15	100.0
Open vs. traditional education on:		
Achievement	26	54.8
Creativity	12	100.0
Self-concept	17	88.2
Attitude toward school	25	92.0
Curiosity	6	100.0
Self-determination	7	85.7
Independence	19	94.7
Freedom from anxiety	8	87.5
Cooperation	6	100.0
Programmed instruction on learning	57	80.7
Adjunct questions on learning:		
After text on recall	98	97.4
After text on transfer	35	74.3
Before text on recall	13	76.9
Before text on transfer	17	29.4
Advance organizers on learning	32	97.0
Analytic revision of instruction on achievement	4	100
Direct instruction on achievement	4	100
Lecture vs. discussion on:		
Achievement	16	68
Retention	7	100
Attitudes	8	00
Student-vs. instructor-centered discussion on:		
Achievement	7	57
Understanding	6	83
Attitude	22	100
Factual vs. conceptual questions on achievement	4	100
Social-psychological climate and learning:		
Cohesiveness	17	0
Satisfaction	17	10
Difficulty	16	0
Formality	17	0
Goal direction	13	7
Democracy	14	0
Environment	15	0
Speed	14	2
Diversity	14	2
Competition	9	0
Friction	17	
Cliqueness	13	
Apathy	13	1
Disorganization	17	
Favoritism	13	
Motivation and learning	232	4
Social class and learning	620	4
Home environment on:		
Verbal achievement	90	1
Math achievement	22	1
Intelligence	20	1
Reading gains	6	1
Ability	8	

Table 5

Correlations and Effect Sizes for Nine Factors in Relation to School Learning

Factor	Number of Studies	Results and Comment
Instruction		
Amount	31	Correlations range from .13 to .71 with a median of .40; partial correlations controlling for ability, socioeconomic status, and other variables range from .09 to .00 with a median of .35
Quality	95	The mean of effect sizes for reinforcement in 39 studies is 1.17, suggesting a 38-point percentile advantage over control groups, although girls and students in special schools might be somewhat more benefited; the mean effect sizes for cues, participation, and corrective feedback in 54 studies is .97, suggesting a 33-point advantage. The mean effect size of similar variables in 18 science studies is .81.
Social-psychological Environment		
Educational	12	On 19 outcomes, social-psychological climate variables added from 1 to 54 (median = 20%) to accountable variance in learning beyond ability and pretests; the signs and magnitudes of the correlations depend on specific scales (see Table 1), level of aggregation (classes and schools higher), nation, and grade level (later grades higher); but not on sample size, subject matter, domain of learning (cognitive, affective, or behavioral), or statistical adjustments for ability and pretests.
Home	18	Correlations of achievement, ability, and motivation with home support and stimulation range from .02 to .62 with a median of .27, multiple correlations range from .29 to .81 with a median of .44; studies of boys and girls and middle-class children in contrast to mixed groups show higher correlations (social classes correlations in 100 studies, by contrast, have a median of .25). The median correlations for three studies of home environment and learning in science is .32.
Media-TV	23	274 correlations of leisure-time television viewing and learning ranged from −.36 to .33 with a median of −.06, although effects appear increasingly deleterious from 10 to 40 hours a week and appear stronger for girls and high-IQ children.
Peer group	10	The median correlation of peer group or friend characteristics such as socioeconomic status and educational aspirations with achievement-test scores, course grades, and educational and occupational aspirations is .24; correlations are higher in urban settings and in studies of students who reported aspirations and achievements of friends. The median of two sciences studies is .24.
Aptitude		
Age-development	9	Correlations between Piaget developmental level and school achievement range from .02 to .71 with a median of .35. The mean correlation in sciences is .40.
Ability	10	From 396 correlations with learning, mean verbal intelligence measures are highest (mean = .72) followed by total ability (.71), nonverbal (.64), and quantitative (.60); correlations with achievement test scores (.70) are higher than those with grades (.57). The mean ability-learning correlation in science is .48.
Motivation	40	Mean correlation with learning is .34, correlations were higher for older samples and for combinations of subjects (mathematics) and measures, but did not depend on type of motivation nor the sex of the samples. The mean of three studies in science is .33.

Table 6

Regressions of Achievement on Productive Factors

Super-Standardized Weights

	Age	Sample Size	Achievement	Attitude	SES	Quality of Instruction	Quantity of Instruction	Education (Class)	Home	Peer	Extra-Curricular Activities	Homework	Stimulation	Media-TV	R^2
Science Achievement	13	2,346		.0111*	.0125*	.0097		.0347**	.0319**	.0069					.31
Science Achievement	17	3,049		.0113*	.0176**	.0101		.0340**	.0113**	.0126**		.0113**			.36
Math Achievement	17	1,480		.0041*	.0051**	.0174** a .0369** b	.0143		.0143					-.0082	.37
Math Attitude	17	1,480	.125			.325** b .100 c			-.225**			.325**	.423**		.09
Social Science Achievement	13	2,426		.0996**	.0325**		.0152**		.0346**						.42
Social Science Attitude	13	2,426	.1470**			.0348** b	.0174*		.0174*					-.0110**	.39
Social Science Achievement	17	2,001		.0586**	.0330**	.0220** a	.0220**		.0256**						.30
Social Science Attitude	17	2,001	.1056**			.0240** b	.0280**		.0217**		.0652**	.0217**	.0186**		.36
			.1250	.0113	.0176	.0248	.0174	.0343	.0098	.0652	.0217	.2218	-.0086		

Note: ~~A key of column numbers will be added.~~ The super-standardized weights are the standardized weights divided by the standard deviations of the dependent variables.

* p<.05
** p<.01
a Traditional instruction
b Student-centered instruction
c Most advanced course

Table 7

Classification of Constructs According to the Model of Educational Productivity

Theorist	Ability	Motivation	Quality of Instruction	Quantity of Instruction	Social Environment of Classroom	Home Environment	Peer Influences	Mass Media
Carroll (1963)	Aptitude Ability to comprehend instructions	Perseverance	Clarity of instruction Matching task to student characteristics	Opportunity to learn (time)				
Cooley and Leinhardt (1975)	General ability Prior achievement	Motivators (internal)	Motivators (external) Structure Instructional Events Attitude toward teachers	Opportunity to learn (time)	Attitudes toward school		Attitudes toward peer	
Bloom (1976)	Prior achievement Reading comprehension Verbal IQ	Attitude toward subject matter Self-concept as learner	Use of cues Reinforcement Feedback and correctives	Participation in learning task (time)	Attitudes toward school			
Fraenkiefeger and Riley (1976)	Pupil background	Intrinsic motivation	Teacher activities	Pupil pursuits (7 time categories)				
Barrett (1978)	Aptitude Prior achievement	Implicit	Clarity of instruction Task difficulty and pacing	Total active learning time Quantity of schooling Time allocated to curriculum activity				
Gagné (1977)	Internal conditions of learning	Implicit	Activating motivation Informing learner of objective Directing attention Stimulating recall Providing learning guidance Enhancing retention Promoting transfer of learning Eliciting performance and providing feedback					
Glaser (1976)	Task learnings already acquired Prerequisite learnings Cognitive style Task specific aptitudes General mediating ability	Implicit	Materials, procedures, and techniques that foster competence (e.g., knowledge structures; learning-to-learn; contingencies of reinforcement) Assessment of effects of instruction					
Skinner (1968)	Task relevant skills	Predispositions	Implanting a predisposition toward learning Structuring knowledge Sequence of materials Specifying rewards and punishments					

Table 8

Effects of Desegregation on Black Achievement

in Three Syntheses

Source	Effect Sizes: Positive Results Percent	Effect Sizes: Mean	Effect Sizes: Standard Deviation	Comments
Krol (1978)	61	.16	.41	Based on 71 comparisons in 55 studies, grade level, mathematics and verbal achievement, and program-duration differences tested and found insignificant.
Crain & Mahard (1982)	62	.10	.25	Percent calculated as sum of 173 positive and half of 50 non-significant comparisons of 321 comparisons in 93 studies; effect-size mean based on 70 studies. With studies as units, significantly larger effects in kindergarten and grade one were found.
"Acceptable Studies"	64	.13	.24	Since the pretest advantage of desegregated groups over control groups was .18, results are calculated for 11 study-weighted means of posttests adjusted for pretests.

Table 9

Inferences from Three Syntheses

About the Effects of Desegregation on Black Achievement

	Percent-Positive Studies		Average Effect Sizes	
	Significance (.05)	Magnitude (67%)	Significance (.05)	Magnitude (.20)
Krol (1978)	?	No	?	No
Crain & Mahard (1982)	?	No	Yes	No
"Acceptable Studies"	No	No	No	No
Conclusion	No?	No	?	No

Note--The criteria for inferences are as follows: The significance required is the standard .05 level calculated for a sign test for a 50-50 split for positive vote counts, and a T test for the difference of the mean effect size from zero, when possible, on independent units of analysis, that is, studies not comparisons. The magnitude criteria are 67 percent of the studies positive and an average effect size of .20, for which the desegregated students would exceed 58 percent of the control-group students.

References

Aitkin, M., Anderson, D., & Hinde, J. Modeling of data on teaching styles (with discussion). Journal of the Royal Statistical Society, Series A, 1981, 144, 419-461.

Aitken, M., Bennett, S. N., & Hesketh, J. Teaching styles and pupil progress: A re-analysis. British Journal of Educational Psychology, 1983, 51, in press.

Bangert, R. L., Kulik, J. A., & Kulik, C.-L. C. Individualized systems of instruction in secondary schools. Ann Arbor: University of Michigan, manuscript, 1981.

Becker, W. C., & Gersten, R. A follow-up of Follow Through. American Educational Research Journal, 1982, 19, 75-92.

Bennett, S. N. Recent research on teaching: A dream, a belief, and a model. British Journal of Educational Psychology, 1978, 48, 127-47.

Bennett, S. N. Teaching styles and pupil progress. London: Open Books, 1976.

Blaug, M. Economic theory in retrospect. New York: Cambridge University Press, 1978.

Bloom, B. S. Human characteristics and school learning. New York: McGraw-Hill, 1976.

Bruner, J. S. Toward a theory of instruction. New York: W. W. Norton & Co., 1966.

Butcher, P. M. An experimental investigation of the effectiveness of a value claim strategy unit for use in teacher education. Sydney, Australia: Macquarie University, unpublished docoral dissertaion, 1981.

Cahen, L. S. Meta-analysis: A technique with promise and problems.

Evaluation in Education, 1980, 4, 37-42.

Carlberg, C., & Kavale, K. The efficacy of special versus regular class placement forexceptional children: A meta-analysis. Journal of Special Education, 1980, 14, 295-309.

Carroll, J. B. A model of school learning. Teachers College Record, 1963, 64, 723-733.

Cohen, P. A. Effectiveness of student-rating feedback for impoving college instruction. Research in Higher Education, 1980, 13, 321- 341.

Cohen, P. A. Student ratings of instruction and student achievement. Review of Educational Research, 1981, 51, 281-309.

Cohen, P. A., Kulik, J. A., & Kulik, C.-L. C. Educational outcomes of tutoring. American Educational Research Journal, 1983, in press.

Cohen, P. A., Ebeling, B. J., & Kulik, J. A. A meta-analysis of outcome studies of visual-based instruction. Education Communication and Technology Journal, 1981, 29, 26-36.

Colosimo, M. L. The effect of practice or beginning teaching on the self concepts and attitudes of teachers: A quantitative synthesis. Chicago: University of Chicago, unpublished doctoral dissertation, 1981.

Cook, T. D., & Campbell, D. T. Quasi-experimentation. Chicago: Rand-McNally, 1979.

Cooley, W. W., & Leinhardt, G. The application of a model for investigating classroom processes. Pittsburgh: University of Pittsburgh Learning Research and Development Center, 1975.

Cooper, H. M. Scientific guidelines for conducting integrative research reviews. Review of Educational Research, 1982, 52, 291-302.

Cooper, H. M., & Rosenthal, R. A comparison of statistical and traditional procedures for summarizing research. Evaluation in

Education, 1980, 4, 33-36.

Crain, R. L., & Mahard, R. E. Desegregation plans that raise black achievement: A review of the research. Santa Monica, Cal.: Rand Corporation, 1982.

Dunkin, M. J. Problems in the accumulation of process-product evidence in classroom research. British Journal of Teacher Education, 1976, 2, 175-187.

Finley, M. J., & Cooper, H. M. The relation between locus of control and academic achievement. Columbia, Missouri: University of Missouri Center for Research in Social Behavior, 1981.

Gagne, R. M. The conditions of learning. Chicago: Holt, Rinehart, & Winston, 1977.

Giaconia, R. M., & Hedges, L. V. Identifying features of open education. Stanford, Calif.: Stanford University, 1982.

Glaser, R. Components of a psychological theory of instruction: Toward a science of design. Review of Educational Research, 1976, 46, 1-24.

Glass, G. V. Integrating findings: The meta-analysis of research. Review of Research in Education, 1977, 5, 351-379.

Glass, G. V., McGaw, B., & Smith, M. L. Meta-analysis of social research. Beverly Hills, Calif.: Sage, 1981.

Graue, M. E., Weinstein, T., & Walberg, H. J. School-based home instruction and learning: A quantitative synthesis. Chicago: University of Illinois, Office of Evaluation Research, 1982.

Graubard, S. R. (Ed.), America's School: Portraits and Perspectives, Daedalus, 1981, 110, 1-175.

Green, J. L. Research on teaching as a linguistic process: A state of the art. Newark: University of Delaware, 1982.

Hanford, B. C., & Hattie, J. A. The relationship between self and

achievement/performance measures. Review of Educational Research, 1982, 52, 123-142.

Harnischfeger, A., & Wiley, D. E. The teaching-learning process in elementary schools: A synoptic view. Curriculum Inquiry, 1976, 6, 5-43.

Haertel, G. D., Walberg, H. J., & Weinstein, T. Psychological models of educational performance: A theoretical synthesis of constructs. Review of Educational Research, 1983, in press.

Hedges, L. V., Giaconia, R. M., & Gage, N. L. Meta-analysis of the effects of open and traditonal instruction. Stanford, Calif.: Stanford University Program on Teaching Effectiveness, 1981.

Horwitz, R. A. Psychological effects of the open classroom. Review of Educational Research, 1979, 49, 71-86.

Jackson, G. B. Methods of integrative reviews. Review of Educational Research, 1980, 50, 438-460.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. Psychological Bulletin, 1981, 89, 47-62.

Krol, R. A. A meta analysis of comparative research on the effects of desegregation on academic achievement. Unpublished doctoral dissertation, Western Michigan University, 1978.

Kulik, C.-L. C., & Kulik, J. A. Effects of ability grouping on secondary school students. Ann Arbor: University of Michigan, manuscript, 1981.

Kulik, C.-L. C., Shwalb, B. J., Kulik, J. A. Programmed instructionn in secondary education. Journal of Educational Research, in press.

Kulik, J. A., Cohen, P. A., & Ebeling, B. J. Effectiveness of programmed instruction in higher education. Educational Evaluation and Policy Analysis, 1980, 2, 51-64.

Kulik, J. A., Kulik, C.-L. C., & Cohen, P. A. Research on audio-tutorial instruction. Research in Higher Education, 1979b, 11, 321-341.

Kulik, J. A., Kulik, C.-L. C., & Cohen, P. A. A meta-analysis of outcome studies of Keller's Personalized System of Instruction. American Psychologist, 1979c, 34, 307-318.

Kulik, J. A., Kulik, C.-L. C. & Cohen, P. A. Effectiveness of computer-based college teaching. Review of Educational Research, 1980, 50, 525-544.

LeCompte, M. D., & Goetz, J. P. Problems of reliability and validity in ethnographic research. Review of Educational Research, 1982, 52, 31-60.

Light, R. J., & Pillemer, D. B. Numbers and narrative: Combining their strengths in research reviews. Harvard Educational Review, 1982, 52, 1-26.

Luiten, J., Ames, W., & Ackerson, G. A meta-analysis of advance organizers on learning and retention. American Educational Research Journal, 1980, 17, 211-218.

Lysakowski, R. S., & Walberg, H. J. Cues, participation, and feedback in instruction: A quantitative synthesis. American Educational Research Journal, 1983, in press.

Ottenbacher, K., & Cooper, H. The effect of class placement on the social adjustment of mentally retarded children. Columbia: University of Missouri Center for Research in Social Behavior, 1981.

Peterson, P. L. Direct instruction reconsidered. In P. L. Peterson & H.

J. Walberg (Eds.), Research on teaching. Berkeley, Calif.: McCutchan, 1979.

Pflaum, S. W., Walberg, H. J., Karegianes, M. L., & Rasher, S. Reading instruction: A quantitative synthesis. Educational Researcher, 1980, 9, 12-18.

Popper, K. R. The logic of scientific discovery. New York: Basic Books, 1959.

Redfield, D. L., & Rousseau, E. W. A meta-analysis of experimental research on teacher questioning behavior. Review of Educational Research, 1981, 51, 237-245.

Rosenthal, R. Combining probabilities and the file drawer problem. Evaluation in education, 1980, 4, 18-21.

Samson, G., Graue, M. E., Weinstein, T., & Walberg, H. J. Academic and occupational performance: A quantitative synthesis. Chicago: University of Illinois Office of Evaluation Research, 1982.

Shulman, L. S., & Tamir, P. Research on teaching in the natural sciences. In R. M. W. Travers (Ed.), Handbook of research on teaching, Second Edition. Chicago: Rand-McNally, 1973.

Slavin, R. E. Cooperative learning. Review of Educational Research, 1980, 50, 315-342.

Smith, M. L. Publication bias and meta-analysis. Evaluation in Education, 1980, 4, 22-24.

Smith, M. L., & Glass, G. V. Meta-analysis of research on class size and its relationship to attitudes. American Educational Research Journal, 1980, 17, 419-433.

Walberg, H. J. A psychological theory of educational productivity. In F. H. Farley & N. Gordon (Eds.), Psychology and Education. Berkeley,

Calif.: McCutchan, 1980.

Walberg, H. J. Education, scientific literacy, and economic productivity. Daedalus, 1983, in press.

Walberg, H. J. What makes schooling effective? Contemporary Education Review 1982, 1, 1-34.

Walberg, H. J., & Haertel, E. H. (Eds.) Research Synthesis: The State of the Art, Evaluation in Education, 1980, 4, 1-142.

Walberg, H. J., & Genova, W. G. School practices and climates that promote integration. Contemporary Educational Psychology, 1983, in press.

Walberg, H. J., Pascarella, E., Haertel, G. D., Junker, L. K., & Boulanger, F. D. Probing a model of educational productivity with national assessment samples of older adolescents. Journal of Educational Psychology, 1982, 74, 295-307.

Walberg, H. J., Schiller, D., & Haertel, G. D. The quiet revolution in educational research. Phi Delta Kappan, 1979, 61 (3), 179-182.

Waxman, H. C., & Walberg, H. J. The relation of teaching and learning. Contemporary Education Review, 1982, 2, 103-120.

Waller, W. The sociology of teaching. New York: Longman's, 1932.

Williams, P. A., Haertel, E. H., Haertel, G. D., & Walberg, H. J. The impact of leisure-time television on school learning. American Educational Research Journal, 1982, 19, 19-50.

Wilkinson. S. S. The relationship of teacher praise and student achievement: A meta-analysis. Gainesville: University of Florida, unpublished doctoral dissertaion, 1980.

Willson, V. L., & Putnam, R. R. A meta-analysis of pretest sensitization effects in experimental design. American Educational Research Journal, 1982, 19, 249-258.